(11)Japanese Patent Application Laid-Open Publication Number Kokai Number: (1997)288631
(43) Kokai Publication Date: November 4, 1997
(54) [Title of the Invention] Fast LAN Control System
(57) [Abstract]
[Problem] Such a problem existed that a copying system is fixed in the case of copying data to between an intra system space buffer and an intra communication adaptor in a communication control system, and hence copying performance declines as a copying target data size changes.
[Solution Means] To select a data copying system (DMA (Direct Memory Access) or PIO (Programmed Input/Output)) between a main memory and a network, corresponding a frame size of transfer data.

[Claims]
[Claim 1] A fast LAN control system in a communication system connected to a plurality of local area networks and transferring data based on an acknowledgement-based highly-reliable network protocol, comprising: a host memory; and a communication adaptor, wherein when copying the data to between said host memory and said communication adaptor, a copying system is selected corresponding to a data length of the data to be copied, and the data is thus copied.
[Claim 2] A fast LAN control system according to Claim 1, wherein a data length serving as a threshold value for controlling the data copy is determined when initializing said communication adaptor.
[Claim 3] A fast LAN control system according to Claim 1, wherein the data length serving as the threshold value for controlling the data copy is set to a fixed value.
[Claim 4] A fast LAN control system according to Claim 1, wherein as to the data length serving as the threshold value for controlling the data copy, periods of data copying time based on a plurality of data copying systems are measured by use of test data when initializing said communication adaptor, and the threshold value for changing the data copying system is controlled based on the measured values.
[Detailed Description of the Invention]
[0001]
[Technical Field of the Invention] The present invention relates to a fast LAN control system, and more particularly to a fast LAN control system suited to copying the data to between a system memory and a memory in a communication adaptor in the case of performing communications based on a highly-reliable protocol in a way that connects to local are networks.
[0002] Two types of systems, a PIO system and a DMA system, are given as systems for transferring the data to between the system memory and the memory in the communication adaptor. PIO is an abbreviation of Programmed Input/Output and is defined as a system for transferring the data by software copy. DMA is an abbreviation of Direct Memory Access and is defined as a system for transferring the data to between an I/O device and a main memory without depending on a CPU.
[0003] The conventional communication control system has, when performing the data communications with another communication control system, a scheme that an adaptor control unit controlling the adaptor copies the data to between the memory within the communication control adaptor and the main memory. At this time, a buffer copy system is fixed to any one of the DMA system and the PIO system, corresponding to the communication adaptor to be connected.
[0004]
[Problems to be Solved by the Invention] FIG. 2 shows a block

diagram of the communication control system.

[0005] A CPU and a host memory are connected to a system bus, and a LAN controller and a memory for the communication adaptor are connected to an I/O bus. A system bus-I/O bus conversion control unit connects the system bus and the I/O bus to each other. Given is an explanation of a case of copying the data to between the host memory and the for-the-communication-adaptor memory of the communication adaptor. To start with, a case of copying the data in the host memory to the memory for the communication adaptor, will be described. In the case of the PIO system, the transmission data in the host memory is temporarily copied to a cache in the CPU and is thereafter copied by the CPU to the memory for the communication adaptor via the system bus-I/O bus conversion control unit. By contrast, the DMA system is that the LAN controller directly accesses the transmission data in the host memory, and copies the data to the for-the-communication-adaptor memory in the communication adaptor. This process is the same with a case of copying the data in the memory for the communication adaptor to the host memory. Namely, a period of copying time is reduced to a degree corresponding to no intermediary of the CPU.

[0006] If the LAN controller has no bus master function, such DMA pre-processing/post-processing is needed as register initialization for the CPU to make the LAN controller conduct the DMA.

[0007] There is a case in which the transfer based on the PIO system is faster depending on a data size to be transferred if a period of actual DMA transfer time is added to the pre-processing/post-processing. Namely, if the transfer data size is small, the DMA pre-processing/post-processing becomes a bottleneck, with the result that the data transfer based on the PIO system becomes faster. If the data size increases, the DMA system gets faster in transfer time than by the PIO system.

[0008] As described above, when transferring the data having a data size "M bytes", a difference occurs in data copying performance depending on the difference in the transfer system, depending on a magnitude of M's value. The data copying performance affects communication performance of the communication control device. In the conventional communication control device, a difference occurs in the communication performance, corresponding to the data size handled by an application running on the communication control device. It is an object of the present invention to change the transfer method, corresponding to the data size of the data transferred to between the main memory and the intra-adaptor memory.

[0009] Means for Solving the Problems] To accomplish the above object, an adaptor control unit in a communication control device includes a data copying control unit. The data copying control

unit is separated into a threshold value determining unit that determines a threshold value for determining a copying system and into a copying control unit. The threshold value determining unit is started up when initializing a communication adaptor. When started up, a fixed value is registered as the threshold value. The copying control unit checks a data size to be copied when transmitted and received, and the copying method is determined by making a comparison with the threshold value in the threshold value determining unit, and the data is thus copied, whereby the object described above is accomplished.

[0010] Further, another means is that the object described is attained by another threshold value setting method using the threshold value determining unit. The threshold value determining unit, when initializing the communication adaptor, measures the DMA pre-processing/post-processing time, and determines the threshold value based on a result of this measurement.

[0011] Still another means is that the object described is attained by still another threshold value setting method using the threshold value determining unit. The threshold value determining unit, when initializing the communication adaptor, generates two types (different in data length) of test data, measures periods of data copy processing time in both of the DMA system and the PIO system with respect to the respective types of test data, and calculates the threshold value from the DMA pre-processing/post-processing time and from the measured value.

[0012]

[Embodiment of the Invention] One embodiment of the present invention will be described with reference to FIGS. 1 - 8.

[0013] FIG. 1 is a diagram of a whole architecture of a communication control system.

[0014] To begin with, each of control blocks will be described.

[0015] A communication control device 1000 is an information device such as a workstation and a personal computer, and connects via a communication adaptor 1001 to a transmission medium 1002 serving to build up a local area network (LAN) such as Ethernet, FDDI (Fibre-Distributed Data Interface), Fast Ethernet and ATM (ATM-LAN)

[0016] A user application program 1003 is a program running within a memory space (user space 1012) in which to operate only a user application in the communication control device. The user application program 1003 transmits and receives the data to and from a protocol control unit 1004 that controls an acknowledgement-based protocol such as TCP/IP (Transmission Control Protocol/Internet Protocol). The protocol control unit 1004 performs the data transmission/reception control according to the acknowledgement-based protocol by use of a transmission buffer 1008 and a reception buffer 1009.

[0017] An adaptor control unit 1006 controls the communication adaptor 1001 and conducts transmission/reception control with respect to a LAN 1002. A network interface control unit 1005 controls interfaces with the adaptor control unit 1006 and with the protocol control unit 1004. The protocol control unit 1004, the network interface control unit 1005, the adaptor control unit 1006 and a communication buffer management control unit are defined as programs running within the system space. The transmission buffer 1008 and the reception buffer 1009 exist in the system space, and are managed by a communication buffer management control unit 1007 by employing an mbuf (memory buffer) data structure. The mbuf data structure will be explained later on.

[0018] The present invention is characterized by obtaining a threshold value for selecting a data copying system at a self-diagnosis time of the adaptor from a data size of the buffer managed based on the mbuf data structure.

[0019] Next, a data copying control unit 1010 in the adaptor control unit 1006 will be described. The data copying control unit 1010 is separated into a threshold value determining unit 1014 and a copy control unit 1015. The threshold value determining unit is started up from an initialization processing unit in the adaptor control unit.

[0020] The threshold value determining unit 1014 has a function of, when initializing the communication adaptor 1001, measuring a length of data copying time based on each of the DMA system and the PIO system between the transmission buffer 1008, the reception buffer 1009 and a buffer 1011 within the communication adaptor, calculating a switching threshold value of the copying system when actually transmitting and receiving the data, and storing the calculated threshold value in a threshold value table provided in the threshold value determining unit 1014.

[0021] Next, transmission/reception buffer control and data copying control in the transmission/reception control unit will be described.

[0022] The transmission/reception data takes an mbuf chain data structure. The transmission/reception process involves automatically setting the data transfer system, the DMA system or the PIO system, between (the buffer managed based on) the mbuf chain data and the transmission/reception buffers in the communication adaptor.

[0023] A scheme of the present invention is that the determination about which system, the DMA system or the PIO system, the data in the buffer managed based on the mbuf data structure is transferred by, is controlled in a way that determines a threshold value for the data size of the buffer managed based on the mbuf data structure.

[0024] Next, the buffer control by a communication buffer management control unit 1007 will be described. The present

embodiment realizes the buffer control by employing the mbuf data structure illustrated in FIG. 3. FIG. 3 illustrates the mbuf management table. An mbuf data structure has fields such as m_next, m_off, m_len, m_type, data and m_act. The mbuf data structure has 128 bytes on the whole, of which 112 bytes are used for storing data.

[0025] The data exists in the data field of the mbuf data structure or in an external page but does not exist simultaneously in both of the locations. An address of the data becomes accessible by adding a value (m_off) in the offset field to an mbuf head address.

[0026] The m_len field represents a byte count of valid data that can be referred to from "offset". An arbitrary length of data is held, and hence a plurality of mbufs (memory buffers) can be linked. This link is established by storing a head address of the next mbuf in the m_next field of the mbuf. The chain data of the linked mbuf is dealt with as the single datagram (packet).

[0027] The m_act field is employed for linking the mbuf chain data to an object list.

[0028] The m_type field is used for registering a data type. FIGS. 4 and 5 each illustrate an example of a structure of a transmission frame using the mbuf. FIG. 4 shows that the m_len field is set to 60 bytes (m_len = 60 bytes), and the m_off field is set to 36 bytes (m_off field = 36 bytes), wherein the 60-byte transmission data is contained in an mbuf structure 3001.

[0029] FIG. 5 shows that an mbuf data structure 5000 is linked to an mbuf data structure 5001, thereby assembling a transmission frame having (64 + 1024) bytes.

[0030] Next, the buffer control and the data transmission process at the transmitting time will be described. The transmission frame is managed based on the mbuf data structure. To start with, the application program 1003 performs the transmission control with respect to the protocol control unit 1004, the transmission frame in the transmission buffer 1008 is structured as illustrated in FIG. 4 or FIG. 5, and the head address of the mbuf data structure is transferred to the adaptor control unit 1006 via the network interface control unit 1005. Next, a processing flow of the data copying control in the adaptor control unit 1006 will be described with reference to FIG. 6. The data copying control unit 1010, at first, acquires the head address of the buffer control data (the mbuf data structure) from which the transmission frame is assembled out of the protocol control unit (6000). Next, as far as the buffer chain exists, the processes, which will hereinafter be explained, are repeated (6001). A data length (m_len) of the mbuf in the mbuf data structure is obtained (6002). A threshold value is obtained from the table in the threshold value determining unit 1014 provided in the data copying control unit 1010 and is compared therewith. If

the data length is larger than the threshold value, the data is transferred based on the DMA system (6004). Whereas if the data length is smaller than the threshold value, the data is transferred based on the PIO system (6005).

[0031] When the data copy is finished, the transmission process of the adaptor is carried out, thereby transmitting the data.

[0032] Next, the buffer control and the data reception process at the receiving time will be described. The adaptor control unit 1006 prepares a buffer for reception, which is managed based on the mbuf data structure, within the system space. When a notification of the reception enters from the communication adaptor 1001, the adaptor control unit 1006 starts up a copying control unit 1015 in the data copying control unit 1010. The copying control unit 1015 refers to the table managed by the communication adaptor 1001 and acquires a reception data size. Then, the reception data size is compared with the threshold value stored in the threshold value table in the threshold value determining unit 1014 provided in the data copying control unit 1010 of the adaptor control unit 1006. If the data length is smaller than the threshold value, the data is transferred based on the PIO system. Whereas if the data length is larger than the threshold value, the data is transferred based on the DMA system.

[0033] An example of a threshold value determining algorithm by the threshold value determining unit 1014 will be next explained. The threshold value determining algorithm is classified into the following three systems.

[0034] 1) Fixed System

The threshold value is fixed. The threshold value is set to 1024 bytes, and the data is copied by the DMA system in the case of the data having a data size larger than 1024 bytes and by the PIO system in the case of being smaller than 1024 bytes.

[0035] 2) DMA Initialization Processing Time Measuring System

A period of DAM initialization processing time is measured, and the threshold value is determined based on this measured value. FIG. 7 shows a processing flow in this case. To begin with, when initializing the communication adaptor (7000), the DAM initialization processing time is measured (7001). The threshold value for selecting the data copying system is calculated based on the measured time. This measuring system is fuzzier than a transfer time diagnosis data generating system, which will be described next, in terms of the threshold value control between the PIO system and the DMA system, but is effective if disabled to copy the test data depending on the specifications of the communication adaptor.

[0036] 3) Transfer Time Diagnosis Data Generating System

Two pieces of test data are generated when initializing the communication adaptor, the data is copied to between the buffer in the system space and the buffer in the communication adaptor,

and the threshold value is calculated from this measured value. Let "M" be a data size, and the threshold value takes such an M's value as to establish the following equation:
(DMA initializing time) + (DMA processing time per byte) x M = (PIO processing time per byte) x M ... (Equation 1)

[0037] Let "A" be a data length of the buffer A, let "B" be a data length of the buffer B, let "D_TA" be data copying time based on the DMA system in the buffer A, let "P_TA" be data copying time based on the PIO system in the buffer A, let "D_TB" be data copying time based on the DMA system in the buffer B, let "P TB" be data copying time based on the PIO system in the buffer B, and the DMA processing time and the PIO processing time are given as follows:

DMA processing time per byte = (A - B)/(D_TA - D_TB)
PIO processing time per byte = (A - B)/(P_TA - P_TB)

[0038] The threshold value M is calculated by measuring D_TA, D_TB, P_TA and P_TB during the DMA initialization when initializing the communication adaptor.

[0039] A processing flow in this case will be explained with reference to FIG. 8.

[0040] The threshold value determining unit is started up when initializing the communication adaptor. The threshold value determining unit, when started up, at first measures the data coping time, and therefore generates two types of buffers (different in data length) in the transmission/reception buffers in the system space and in the communication adaptor (8001). Then, with respect to the two types of buffers, the data is copied to between the intra system space buffer and the intra communication adaptor buffer by employing the both of the DMA system and the PIO system (8002).

[0041] During the DMA initialization, periods of data copying time with respect to the buffers having the data sizes A and B on the basis of both of the DMA system and the PIO system, are measured (8003). The threshold value is calculated from the Equation (1) (8004). Then, the threshold value is registered in the threshold value table, thus finishing the process by the threshold value determining unit.

[0042] As to the buffer control, the data copying system for the buffer control using the mbuf data structure has been described, however, the data copying system according to the present invention is effective also in the case of the buffer control that does not use the data link model as by the mbuf data structure.

[0043] According to the present invention, the data copying system is automatically selected corresponding to the data size for copying, thereby improving the copying performance of the frame data.

[0044] Thus, the threshold value based on the data length for copying is set, and the data copying system is controlled, which

prevents the data copying time from linearly increasing with respect to the data length, thereby reducing the data copy processing time that is a bottleneck to the data communications.

[0045] A difference in the data copying performance affects the communication performance of the communication control device. The communication control device performing the data copy based on only the DMA system has an effect in the case of communicating the data having a large data size as by FTP (File Transfer Protocol), but has none of the effect in the communication performance in the case of communicating the data having a small data size as by Telnet. In contrast, the communication control device performing the data copy based on only the PIO system is effective in the application that handles a small data size. The present invention utilizes the data copy based on the PIO system in the case of the small data size to be handled and the data copy based on the DMA system in the case of the large data size to be handled, whereby the fast data transfer can be done irrespective of the data size to be handled by the communication application.

[Brief Description of the Drawings]

[FIG. 1] A diagram of a whole architecture of a communication control system.

[FIG. 2] A block diagram of the communication control system.

[FIG. 3] A diagram of a mbuf buffer management table.

[FIG. 4] A diagram of an example 1 of a structure of the mbuf buffer management table.

[FIG. 5] A diagram of an example 2 of the structure of the mbuf buffer management table.

[FIG. 6] A diagram showing a flow of a buffer copying process.

[FIG. 7] A diagram showing a processing flow 1 of a threshold value determining process.

[FIG. 8] A diagram showing a processing flow 2 of the threshold value determining process.

[Description of Numerals and Symbols]

1000... communication control device, 1001... communication adaptor, 1002... LAN, 1003... user application, 1004... protocol control unit, 1005... network interface control unit, 1006... adaptor control unit, 1007... communication buffer management control unit, 1008... transmission buffer, 1009... reception buffer, 1010... data copying control unit, 1011...intra adaptor buffer, 1012... user space, 1013... system space, 2000... CPU, 2001...host memory, 2002... system bus, 2003...system bus I/O bus conversion control unit, 2004... I/O bus, 2005... LAN controller, 2006...memory for communication adaptor

FIG. 1:
1000... COMMUNICATION CONTROL DEVICE,
1003... USER APPLICATION PROGRAM,
1004... PROTOCOL CONTROL UNIT,
1009... RECEPTION BUFFER,
1008... TRANSMISSION BUFFER,
1007... COMMUNICATION BUFFER MANAGEMENT CONTROL UNIT,
1005... NETWORK INTERFACE CONTROL UNIT,
1006... ADAPTOR CONTROL UNIT,
1010... DATA COPYING CONTROL UNIT,
1014... THRESHOLD VALUE DETERMINING UNIT,
1015... COPYING CONTROL UNIT,
1001... COMMUNICATION ADAPTOR,
1011... ADAPTOR BUFFER,

FIG. 2:
2005... LAN CONTROLLER,
2006... MEMORY FOR COMMUNICATION ADAPTOR,

FIG. 3:
A... 112 BYTES,

FIG. 4:
A... 36 BYTES,
B... 60 BYTES,

FIG. 5:
A... 36 BYTES,
B... 64 BYTES,
C... 1000 BYTES,
D... 1024 BYTES,

FIG. 6:
6000... ACQUIRE HEAD ADDRESS OF BUFFER CONTROL DATA ASSEMBLING TRANSMISSION FRAME FROM HIGH-ORDER CONTROL UNIT,
6001... BUFFER CHAIN EXISTS,
6002... OBTAINS mbuf DATA LENGTH,
6003... COMPARE WITH THRESHOLD VALUE,
A... LARGER THAN THRESHOLD VALUE,
B... SMALLER THAN THRESHOLD VALUE,
6004... DATA TRANSFER BASED ON DMA SYSTEM,
6005... DATA TRANSFER BASED ON PIO SYSTEM,

FIG. 7:
7000... ADAPTOR INITIALIZING PROCESS,
7001... MEASURE PRE-PROCESSING/POST-PROCESSING TIME WITH DMA.
7002... DETERMINE THRESHOLD VALUE FROM DMA PRE-PROCESSING/POST-PROCESSING.

FIG. 8:
8000... ADAPTOR INITIALIZING PROCESS,
8001... GENERATE TWO TYPES OF BUFFERS FOR MEASURING DATA COPYING TIME,
8002... COPY DATA BY USE OF DMA, PIO FOR TOW TYPES OF BUFFERS,
8003... GENERATE LINEAR EQUATION FOR DATA SIZE AND TRANSFER TIME ABOUT DMA, PIO,
8004... SOLVE SIMULTANEOUS EQUATIONS OF DMA AND PIO,
8005... REGISTER DATA SIZE AT THAT TIME AS THRESHOLD VALUE,

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 09-288631
(43)Date of publication of application : 04.11.1997

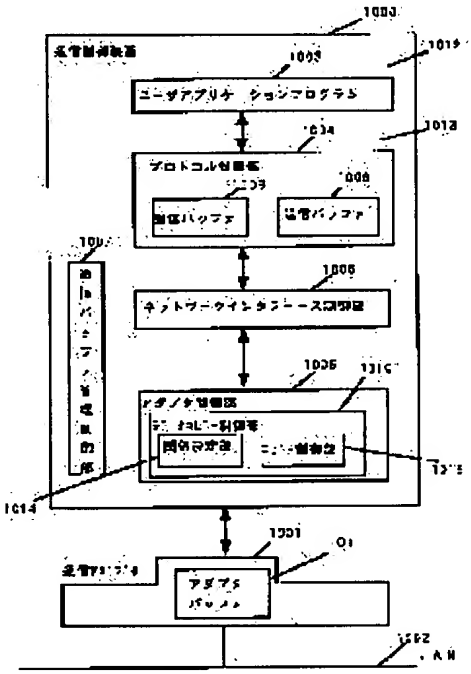
(51)Int.Cl. G06F 13/00
G06F 13/28

(21)Application number : 08-102221
(22)Date of filing : 24.04.1996

(71)Applicant : HITACHI LTD
(72)Inventor : HIGUCHI HIDEMITSU
KOBAYASHI ICHIJI
HOTTA MASAYA
ONO SHUJI

(54) FAST LAN CONTROL SYSTEM

(57)Abstract:
PROBLEM TO BE SOLVED: To improve the copy performance of frame data by selecting a copy system according to the length of data to be copied between a host memory and a communication adapter.
SOLUTION: An adapter control part 1006 prepares a receiving buffer which is managed by an mbuf data structure in a system space. When the receiving communication is inputted from a communication adapter 1001, the part 1006 starts a copy control part 1015 included in a data copy control part 1010. The part 1015 refers to a table which is managed by the adapter 1001 to know the receiving data length. This data length is compared with the threshold contained in a threshold table included in a threshold decision part 1014 which is included in the part 1015. If the data length is smaller than the threshold, the transfer of data is carried out by a program input/output system. If the data length is larger than the threshold, the transfer of data is carried out by a direct memory access system.



LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

(19)日本国特許庁（ＪＰ）　(12) **公開特許公報（Ａ）**　(11)特許出願公開番号

特開平9－288631

(43)公開日　平成９年(1997)11月４日

(51)Int.Cl.[6]		識別記号	庁内整理番号	Ｆ Ｉ			技術表示箇所
Ｇ０６Ｆ	13/00	３５３		Ｇ０６Ｆ	13/00	３５３Ｎ	
	13/28	３３０			13/28	３３０	

審査請求　未請求　請求項の数４　ＯＬ　（全 ７ 頁）

(21)出願番号　特願平8－102221

(22)出願日　平成８年(1996)４月24日

(71)出願人　000005108
株式会社日立製作所
東京都千代田区神田駿河台四丁目６番地

(72)発明者　樋口　秀光
神奈川県川崎市幸区鹿島田890番地の12株式会社日立製作所情報・通信開発本部内

(72)発明者　小林　一司
神奈川県海老名市下今泉810番地株式会社日立製作所オフィスシステム事業部内

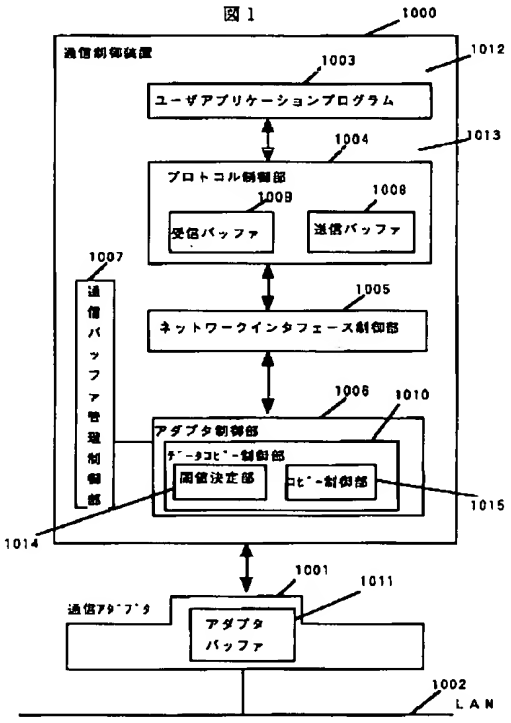
(74)代理人　弁理士　小川　勝男

最終頁に続く

(54)【発明の名称】　高速ＬＡＮ制御方式

(57)【要約】
【課題】通信制御システム内におけるシステム空間内バッファと通信アダプタ内バッファ内のデータコピーにおいて、コピー方式を固定していたため、コピーするデータ量の変化に対して、コピー性能が劣化するという課題があった。
【解決手段】転送データのフレームサイズに応じて、主メモリとネットワーク間のデータコピー方式（ＤＭＡ又は、ＰＩＯ）を選択する。

図１



1

【特許請求の範囲】

【請求項１】複数のローカルエリアネットワークに接続し、送達確認型の高信頼なネットワークプロトコルでデータ転送を行う通信システムにおいて、ホストメモリと通信アダプタの間でデータコピーをする際にコピーするデータのデータ長に応じてコピー方式を選択し、コピーすることを特徴とする高速ＬＡＮ制御方式。

【請求項２】請求項１に記載の高速ＬＡＮ制御方式において、データコピーを制御する閾値となるデータ長を通信アダプタの初期化時に決定することを特徴とする高速ＬＡＮ制御方式。

【請求項３】請求項１に記載の高速ＬＡＮ制御方式において、データコピーを制御する閾値となるデータ長を固定値とすることを特徴とする高速ＬＡＮ制御方式。

【請求項４】請求項１に記載の高速ＬＡＮ制御方式において、データコピーを制御する閾値となるデータ長を通信アダプタの初期化時にテストデータを使って複数のデータコピー方式によるデータコピー時間を測定し、その測定値からデータコピー方式を変更する閾値を制御することを特徴とする高速ＬＡＮ制御方式。

【発明の詳細な説明】

【０００１】

【発明の属する技術分野】本発明は、高速ＬＡＮ制御方式に関し、特にローカルエリアネットワークに接続し、高信頼のプロトコルを用いて通信を行う場合にシステムメモリと通信アダプタ内メモリ間のデータコピーに好適な高速ＬＡＮ制御方式に関する。

【０００２】

【従来の技術】システムメモリと通信アダプタ内メモリ間のデータ転送方式には、ＰＩＯとＤＭＡの２方式がある．ＰＩＯとは、ＰｒｏｇｒａｍＩｎｐｕｔ／Ｏｕｔｐｕｔの略で、ソフトウエアコピーによりデータ転送を行う方式である。ＤＭＡはＤｉｒｅｃｔ　ＭｅｍｏｒｙＡｃｃｅｓｓの略でＣＰＵによらないＩ／Ｏ機器と主メモリ間のデータ転送方式である。

【０００３】従来の通信制御システムでは、他の通信制御システムとデータ通信を行う際、アダプタを制御するアダプタ制御部が、通信制御アダプタ内のメモリと主メモリ間のデータコピーを行っている。このとき、接続する通信アダプタに応じてバッファコピー方式をＤＭＡかＰＩＯのどちらか一方に固定していた。

【０００４】

【発明が解決しようとする課題】図２は、通信制御システムをブロック図で示している。

【０００５】システムバス上にＣＰＵとホストメモリが接続され、Ｉ／Ｏバス上にＬＡＮコントローラと通信アダプタ用メモリが接続されている。システムバスとＩ／Ｏバスの間は、システムバスーＩ／Ｏバス変換制御部で接続されている。ホストメモリ内のデータと通信アダプタの通信アダプタ用メモリとの間でデータをコピーする

2

場合について説明する。まず、ホストメモリ内のデータを通信アダプタ用メモリにコピーする場合について説明する。ＰＩＯ方式の場合ホストメモリ内の送信データは、ＣＰＵ内のキャッシュに一時コピーされ、しかる後、ＣＰＵによってシステムバスーＩ／Ｏバス変換制御部を通して通信アダプタ用メモリにコピーされる。これに対し、ＤＭＡ方式は、ホストメモリ内の送信データをＬＡＮコントローラが直接アクセスし、通信アダプタ内の通信アダプタ用メモリにコピーする。通信アダプタ用メモリ内データをホストメモリにコピーする場合も同様である。つまり、ＣＰＵを介さない分、コピー時間は速くなる。

【０００６】しかし、ＬＡＮコントローラにバスマスタ機能がないとＣＰＵがＬＡＮコントローラに対し、ＤＭＡを行なわせるためのレジスタ初期化等のＤＭＡ前処理・後処理が必要になる。

【０００７】この前・後処理に実際のＤＭＡ転送時間を加えると転送するデータ量によってＰＩＯ方式による転送の方が速い場合がある。つまり、転送データ量が小さい場合、ＤＭＡの前後処理時間がネックになってＰＩＯ方式によるデータ転送の方が速くなる。データ量が大きくなるとＤＭＡ方式による転送時間の方がＰＩＯ方式よりも速くなる。

【０００８】以上のように、データ量Ｍバイトのデータを転送する場合、Ｍの大きさによって転送方式の差によりデータコピー性能に差が生じる。データコピー性能は、通信制御装置の通信性能に影響を与える。従来の通信制御装置では通信制御装置上で動作するアプリケーションの扱うデータ量に応じて通信性能に差が生じていた本発明の目的は、主メモリとアダプタ内のメモリ間データ転送を転送するデータ量に応じて、転送方法を変えることである。

【０００９】

【課題を解決するための手段】前記目的を達成させるため、通信制御装置内のアダプタ制御部にデータコピー制御部を設ける。データコピー制御部は、コピー方式を決定する閾値を決める閾値決定部とコピー制御部に分かれる。閾値決定部は、通信アダプタの初期化時に起動される。起動されると固定値を閾値として登録する。コピー制御部が送受信時にコピーするデータ量を調べ、閾値決定部内の閾値と比較し、コピー方法を決定してデータコピーをすることにより前記目的は達成される。

【００１０】また、別の手段として閾値決定部による別の閾値設定方法によって前記目的は達成される。閾値決定部は通信アダプタ初期化時にＤＭＡ前後処理時間を測定し、その結果を基に閾値を決定する。

【００１１】また、別の手段として閾値決定部による閾値設定方法によって前記目的は、達成される。閾値決定部は、通信アダプタ初期化時に２種類（データ長の異なる）のテストデータを作成し、それぞれのデータに対

し、ＤＭＡとＰＩＯの両方式でデータコピー処理時間を測定し、ＤＭＡ前後処理時間と測定値から閾値を計算する。

【００１２】

【発明の実施の形態】本発明の一実施例を図１～図８を使って説明する。

【００１３】図１は、通信制御システムの全体構成図である。

【００１４】まず、各制御ブロックについて説明する。

【００１５】通信制御装置１０００は、ワークステーション、パーソナルコンピュータ等の情報機器であり、例えば、Ethernet、 FDDI、 FastEthernet、 ATMなどのようなローカルエリアネットワーク（ＬＡＮ）を構築するような伝送媒体１００２に通信アダプタ１００１を介して接続し、他の情報機器とデータ転送を行う。

【００１６】ユーザアプリケーションプログラム１００３は、通信制御装置内のユーザアプリケーションだけが動作するメモリ空間（ユーザ空間１０１２）内で動作するプログラムである。ユーザアプリケーションプログラム１００３は、ＴＣＰ／ＩＰのような送達確認型のプロトコル制御を行うプロトコル制御部１００４に対して、データの送受信を行う。プロトコル制御部１００４は、送信バッファ１００８及び受信バッファ１００９を使って送達確認型プロトコルに従ったデータ送受信制御を行う。

【００１７】アダプタ制御部１００６は、通信アダプタ１００１を制御し、ＬＡＮ１００２に対して送受信制御を行う。ネットワークインタフェース制御部１００５は、アダプタ制御１００６とプロトコル制御部１００４とのインタフェース制御を行う。プロトコル制御部１００４、ネットワークインタフェース制御部１００５、アダプタ制御部１００６、通信バッファ管理制御部は、システム空間内で動作するプログラムである。送信バッファ１００８と受信バッファ１００９は、システム空間内にあり、ｍｂｕｆデータ構造体を使って通信バッファ管理制御部１００７に管理されている。ｍｂｕｆデータ構造体については、後に説明する。

【００１８】本発明では、アダプタの自己診断時にデータコピー方式を選択するための閾値をｍｂｕｆデータ構造体により管理されるバッファのデータ量から求めることを特徴とする。

【００１９】つぎに、アダプタ制御部１００６内のデータコピー制御部１０１０について説明する。データコピー制御部１０１０は、閾値決定部１０１４とコピー制御部１０１５に分けられる。閾値決定部は、アダプタ制御部内の初期処理部から起動される。

【００２０】閾値決定部１０１４は、通信アダプタ１００１の初期化時に送受信バッファ１００８、１００９と通信アダプタ内バッファ１０１１との間のＤＭＡ及びＰＩＯによるデータコピー時間を測定し、実際のデータ送受信時のコピー方式の切替え閾値を算出し、閾値決定部１０１４内にある閾値テーブルに格納する機能を持つ。

【００２１】つぎに、送受信制御部における送受信バッファ制御とデータコピー制御について説明する。

【００２２】送受信データは、ｍｂｕｆチェインデータ構造を取っている。送受信処理では、このｍｂｕｆチェインデータと通信アダプタ内の送受信バッファとの間のデータ転送方式をＤＭＡかＰＩＯのどちらかに自動設定する。

【００２３】本発明では、ｍｂｕｆデータ構造体によって管理されるバッファ内のデータ転送をＤＭＡで行うかＰＩＯで行うかの判定をｍｂｕｆデータ構造体によって管理されるバッファのデータ量に閾値を決めることにより制御する。

【００２４】つぎに通信バッファ管理制御部１００７によるバッファ制御について説明する。　本実施例では、バッファ制御は、図３に示すｍｂｕｆデータ構造体を使って実現する。ｍｂｕｆ管理テーブルを図３に示す。ｍｂｕｆのフィールドは、m_next、 m_off、 m_len、 m_type、 data、 m_actで構成され、全体で１２８バイト、そのうち、１１２バイトがデータを蓄えるための部分である。

【００２５】データは、ｍｂｕｆ内部のデータ領域か外部のページに存在するが両方同時ではない。データのアドレスは、ｍｂｕｆの先頭アドレスにオフセットフィールドの値(m_off)を足すことでアクセスできる。

【００２６】m_lenはオフセットから参照できる有効なデータのバイト数を示す。任意の長さのデータを保有するため、複数のｍｂｕｆをリンクする事ができる。このリンクは、ｍｂｕｆのｍ＿ｎｅｘｔフィールドにつぎのｍｂｕｆの先頭アドレスを格納することにより実現される。リンクされたｍｂｕｆのチェインは、単一のデータ（パケット）として扱われる。

【００２７】ｍ＿ａｃｔフィールドはｍｂｕｆチェインをオブジェクトリストにリンクするために使用する。

【００２８】ｍ＿ｔｙｐｅフィールドは、データのタイプ種別を登録するために使用される。図４、図５にｍｂｕｆを使った送信フレームの構成例を示す。図４では、ｍ＿ｌｅｎ＝６０バイト、ｍ＿ｏｆｆ＝３６バイトに設定されており、６０バイトの送信データが、ｍｂｕｆデータ構造体３００１の中に入っていることを示している。

【００２９】図５では、ｍｂｕｆデータ構造体５０００とｍｂｕｆデータ構造体５００１がリンクして６４＋１０２４バイトの送信フレームを構成していることを示している。

【００３０】つぎに送信時のバッファ制御とデータ送信処理について説明する。送信フレームは、ｍｂｕｆデータ構造体によって管理されている。まず、アプリケーションプログラム１００３が、プロトコル制御部１００４

に対して送信制御を行うと、送信バッファ１００８内の送信フレームは、図４又は図５のように構成され、ｍｂｕｆデータ構造体の先頭アドレスがネットワークインタフェース制御部１００５を通してアダプタ制御部１００６に渡される。次にアダプタ制御部１００６内のデータコピー制御部１０１０によるデータコピー制御の処理の流れについて図６を使って説明する。データコピー制御部１０１０は、まず、プロトコル制御部から送信フレームを構成するバッファ制御データ（ｍｂｕｆデータ構造体）の先頭アドレスを獲得する（６０００）。次に、バッファチェインがある限り、以下に説明する処理を繰り返す（６００１）。ｍｂｕｆデータ構造体中のｍｂｕｆのデータ長（ｍ＿ｌｅｎ）を求める（６００２）。データコピー制御部内にある閾値決定部内のテーブルより閾値を求め、比較する。データ長が閾値より大きい場合は、ＤＭＡ方式によるデータ転送を行う（６００４）。データ長が閾値より小さい場合は、ＰＩＯ方式によるデータ転送を行う（６００５）。

【００３１】データコピーが終了するとアダプタの送信処理を行い、データが送信される。

【００３２】つぎに、受信時のバッファ制御とデータ受信処理について説明する。アダプタ制御部１００１は、システム空間内にｍｂｕｆデータ構造体で管理される受信用のバッファを用意する。受信通知が通信アダプタ１００１から入るとアダプタ制御部１００６は、データコピー制御部１０１０内のコピー制御部１０１５を起動する。コピー制御部１０１５は、通信アダプタが管理するテーブルを参照し、受信データサイズを得る。そして、受信データサイズとアダプタ制御部１００６のデータコピー制御部１０１０内にある閾値決定部１０１４内の閾値テーブルに格納されている閾値と比較し、データ長が閾値より小さい場合は、ＰＩＯ方式によるデータ転送を行う。データ長が閾値より大きい場合は、ＤＭＡ方式によるデータ転送を行う。

【００３３】閾値決定部による閾値決定アルゴリズムの例について次に説明する。閾値決定アルゴリズムには、次の３方式がある。

【００３４】１）固定方式

閾値を固定にする。閾値を１０２４バイトとし、１０２４バイトより大きいデータの場合は、ＤＭＡ、小さいデータの場合は、ＰＩＯ方式でデータコピーを行う。

【００３５】２）ＤＭＡ初期処理時間測定方式

ＤＭＡ初期処理時間を測定し、その値によって閾値を決定する。この場合の処理の流れを図７に示す。まず、通信アダプタの初期化時（７０００）にＤＭＡ初期化処理時間を測定する（７００１）。測定時間を基にデータコピー方式の選択の閾値を算出する。この方式は、つぎに説明する転送時間診断データ作成方式に比べ、ＰＩＯ方式とＤＭＡ方式の間の閾値制御があいまいであるが、通信アダプタの仕様によりテストデータのコピーができない場合有効である。

【００３６】３）転送時間診断データ作成方式

通信アダプタの初期化時に２つのテスト用データを作成し、システム空間と通信アダプタ内のバッファとの間でデータコピーを行い、その測定値により閾値を算出する。データ量をＭとすると、

（ＤＭＡ初期化時間）＋（１ﾊﾞｲﾄ当たりのDMA処理時間）×Ｍ=（１ﾊﾞｲﾄ当たりのＰＩＯ処理時間）×Ｍ（式１）

となるようなＭの値が閾値となる。

【００３７】バッファＡのデータ長＝Ａ、バッファＢのデータ長＝Ｂとし、バッファＡのＤＭＡ方式によるデータコピー時間をＤ＿ＴＡ、ＰＩＯ方式によるデータコピー時間をＰ＿ＴＡとし、バッファＢのＤＭＡ方式によるデータコピー時間をＰ＿ＴＢ、ＰＩＯ方式によるデータコピー時間をＰ＿ＴＢとすると、

１ﾊﾞｲﾄ当たりのDMA処理時間　＝　(A -B)/(D_TA - D_TB)

１ﾊﾞｲﾄ当たりのＰＩＯ処理時間　＝　(A -B)/(P_TA - P_TB)　となる。

【００３８】通信アダプタ初期化時にＤＭＡ初期化時間、D_TA、 D_TB、 P_TA、 P_TBを測定することにより、閾値Ｍが算出される。

【００３９】この場合の処理の流れを図８を使って説明する。

【００４０】通信アダプタ初期化時に閾値決定部を起動する。閾値決定部は起動されると、まず、データコピー時間を測定するため、システム空間と通信アダプタ内送受信バッファにバッファを２種類作成する（データ長の異なるもの）（８００１）。そして、２種類のバッファに対し、ＤＭＡ、ＰＩＯの両方式を使って、システム空間内バッファと通信アダプタ内バッファの間でデータコピーををする（８００２）。

【００４１】ＤＭＡ初期化時間、データ量Ａ、Ｂのバッファに対するＤＭＡ、ＰＩＯ両方式によるデータコピー時間を測定する（８００３）。式１から閾値を計算する（８００４）。そして、閾値を閾値テーブルに登録し、閾値決定部の処理は終了する。

【００４２】バッファ制御について、ｍｂｕｆデータ構造体を使ったバッファ制御に対するデータコピー方式を説明したが、ｍｂｕｆデータ構造体のようなデータリンク形式を使わないバッファ制御の場合に対しても本発明によるデータコピー方式は、有効である。

【００４３】

【発明の効果】本発明では、データコピー方式をコピーするデータ量に応じて自動選択することにより、フレームデータのコピー性能が向上する。

【００４４】このように、コピーするデータ長による閾値を設定し、データコピー方式を制御することは、データコピー時間がデータ長に対して、線形的に増加するのを防ぐため、データ通信のボトルネックになっているデ

ータコピー処理時間を軽減する。

【００４５】データコピー性能の差は通信制御装置の通信性能に影響を与える。ＤＭＡ方式のみのデータコピーを行っている通信制御装置はＦＴＰのような大量のデータを通信する場合に効果があるが、ｔｅｌｎｅｔのようなデータ量の小さい通信の場合通信性能に対する効果は見られない。反対にＰＩＯ方式のみのデータコピーを行っている通信制御装置は、扱うデータ量の小さいアプリケーションに対して効果がある。本発明では扱うデータ量が小さい場合は、ＰＩＯ方式によるデータコピーを使用し、扱うデータ量が大きい場合は、ＤＭＡ方式によるデータコピーを使用することにより、通信アプリケーションの扱うデータ量に関係なく高速データ転送が行える。

【図面の簡単な説明】

【図１】通信制御システム全体構成図である。

【図２】通信制御システムのブロック図である。

【図３】mbufバッファ管理テーブルである。

【図４】ｍｂｕｆバッファ管理テーブルの構成例１である＊る。

【図５】ｍｂｕｆバッファ管理テーブルの構成例２である。

【図６】バッファコピーの処理の流れである。

【図７】閾値決定処理の処理の流れ１である。

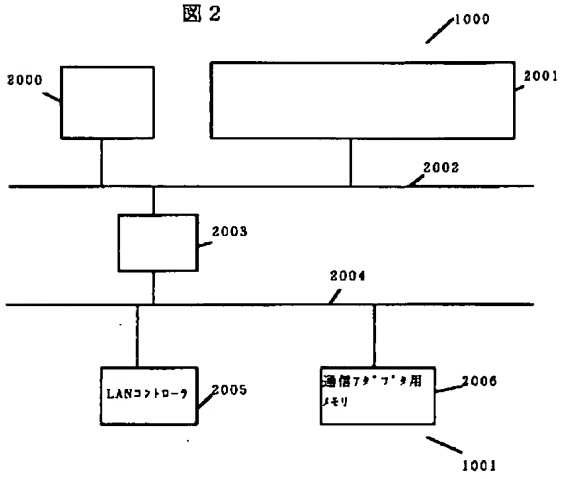
【図８】閾値決定処理の処理の流れ２である。

【符号の説明】

１０００…通信制御装置、　１００１…通信アダプタ、１００２…ＬＡＮ、１００３…ユーザアプリケーション、１００４…プロトコル制御部、１００５…ネットワークインタフェース制御部、１００６…アダプタ制御部、１００７…通信バッファ管理制御部、１００８…送信バッファ、　１００９…受信バッファ、１０１０…データコピー制御部、１０１１…アダプタ内バッファ、１０１２…ユーザ空間、　１０１３…システム空間、　２０００…ＣＰＵ、２００１…ホストメモリ、２００２…システムバス、２００３…システムバスーＩ／Ｏバス変換制御部、２００４…Ｉ／Ｏバス、　２００５…ＬＡＮコントローラ、２００６…通信アダプタ用メモリ。

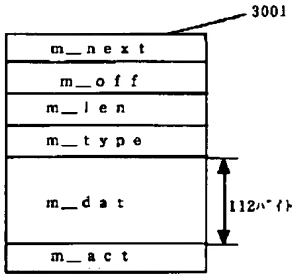
【図２】

図２



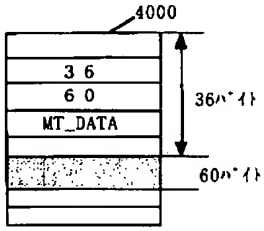
【図３】

図３



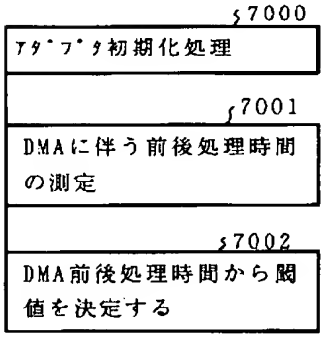
【図４】

図４



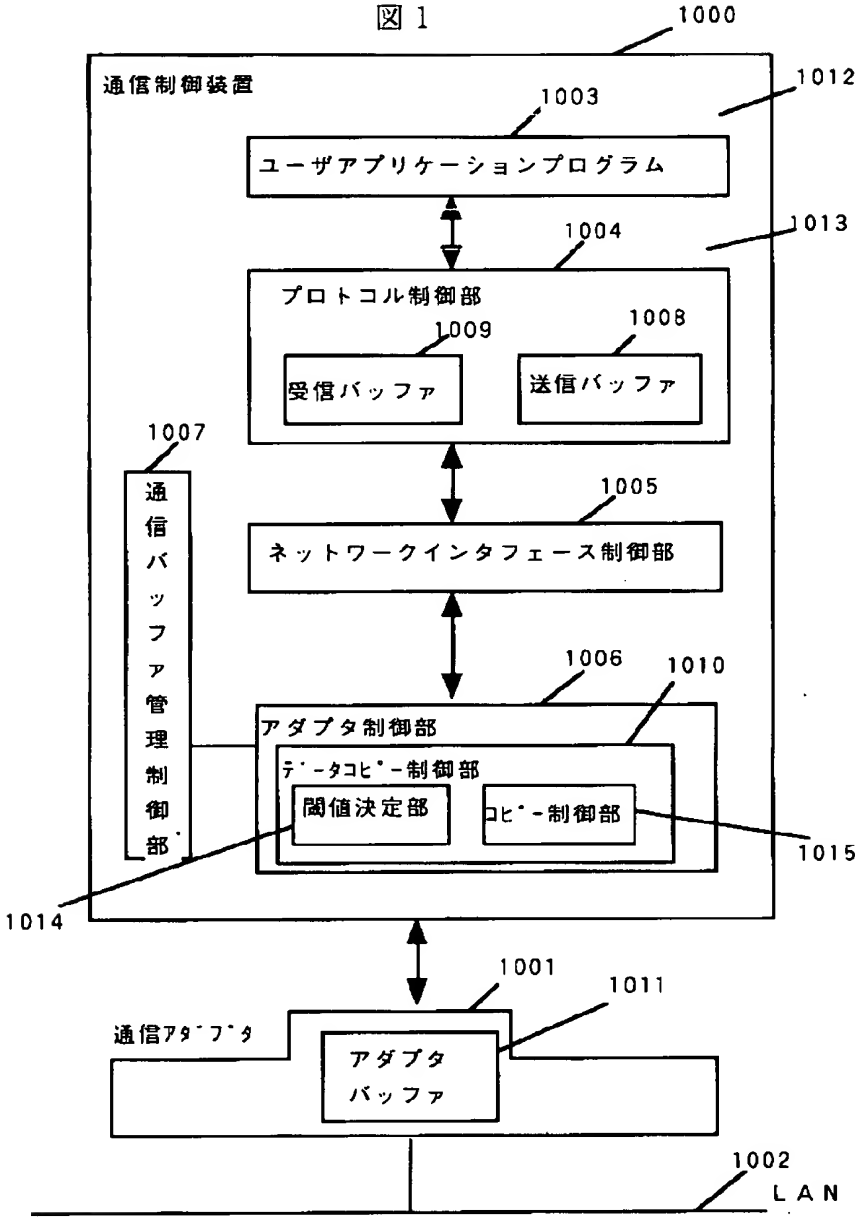
【図７】

図７



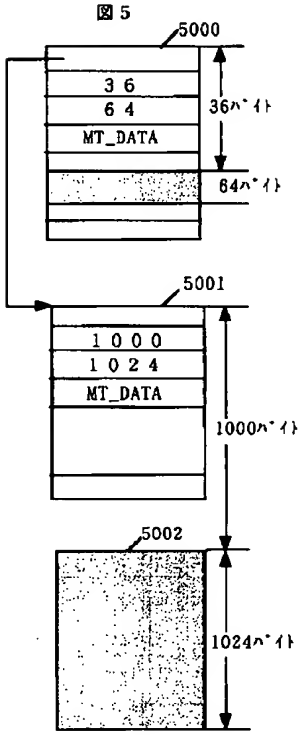
【図１】

図１


1000
通信制御装置
1012
1003
ユーザアプリケーションプログラム
1013
1004
プロトコル制御部
1009
1008
受信バッファ
送信バッファ
1007
通信バッファ管理制御部
1005
ネットワークインタフェース制御部
1006
1010
アダプタ制御部
ﾃﾞｰﾀｺﾋﾟｰ制御部
閾値決定部
ｺﾋﾟｰ制御部
1015
1014
1001
1011
通信ｱﾀﾞﾌﾟﾀ
アダプタバッファ
1002
LAN

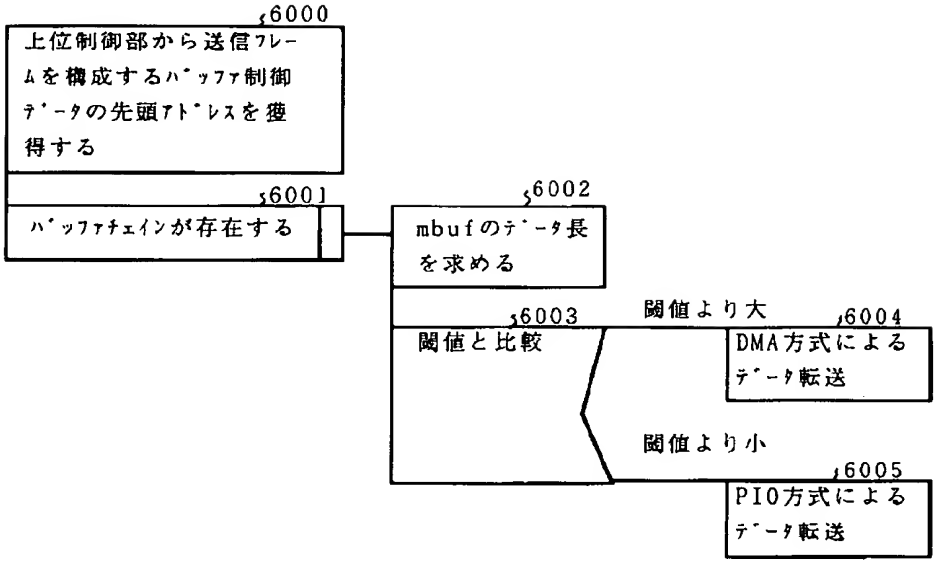
【図５】

図５


5000
３６
６４
MT_DATA
36ﾊﾞｲﾄ
64ﾊﾞｲﾄ
5001
１０００
１０２４
MT_DATA
1000ﾊﾞｲﾄ
5002
1024ﾊﾞｲﾄ

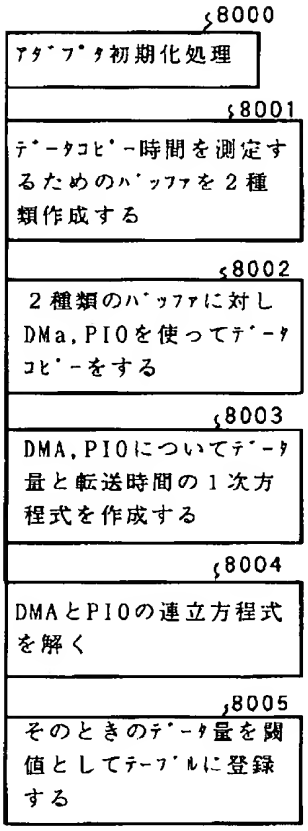
【図６】

図６



【図８】

図８



フロントページの続き

(72)発明者　堀田　匡哉
神奈川県横浜市中区尾上町６丁目81番地日立ソフトウェアエンジニアリング株式会社内

(72)発明者　大野　修司
神奈川県川崎市幸区鹿島田890番地の12株式会社日立製作所情報・通信開発本部内